
PATENT
Attorney Docket No.: 16869B-080700US
Client Ref. No.: HAL273
(340300834US01)

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

| | |
|---|---|
| In re application of:<br>KENJI YAMAGAMI<br>Application No.: 10/758,971<br>Filed: January 15, 2004<br>For: DISTRIBUTED REMOTE COPY SYSTEM<br>Customer No.: 20350 | Examiner: Unassigned<br>Technology Center/Art Unit: 2131<br>Confirmation No.: 6513<br>**PETITION TO MAKE SPECIAL FOR NEW APPLICATION UNDER M.P.E.P. § 708.02, VIII & 37 C.F.R. § 1.102(d)** |

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This is a petition to make special the above-identified application under MPEP § 708.02, VIII & 37 C.F.R. § 1.102(d). The application has not received any examination by an Examiner.

(a) The Commissioner is authorized to charge the petition fee of $130 under 37 C.F.R. § 1.17(i) and any other fees associated with this paper to Deposit Account 20-1430.

(b) All the claims are believed to be directed to a single invention. If the Office determines that all the claims presented are not obviously directed to a single invention, then Applicants will make an election without traverse as a prerequisite to the grant of special status.

10/12/2005 HLE333 00000087 201430 10758971
01 FC:1464 130.00 DA

(c) Pre-examination searches were made of U.S. issued patents, including a classification search and a key word search. The classification search was conducted on or around June 15, 2005 covering Class 707 (subclasses 10 and 200-204), Class 711 (subclasses 112-114, 167, and 168), and Class 714 (subclasses 1, 6, 13, 20, and 39), by a professional search firm, Mattingly, Stanger, Malur & Brundidge, P.C. The key word search was performed on the USPTO full-text database including published U.S. patent applications. A search for foreign art was also conducted using the European Patent Office's ESPACENET database and Japanese patent database.

(d) The following references, copies of which are attached herewith, are deemed most closely related to the subject matter encompassed by the claims:

(1) U.S. Patent No. 5,446,871;

(2) U.S. Patent No. 5,504,861;

(3) U.S. Patent No. 5,734,818;

(4) U.S. Patent No. 6,658,540 B1;

(5) U.S. Patent Publication No. 2003/0014432 A1;

(6) U.S. Patent Publication No. 2003/0074378 A1;

(7) U.S. Patent Publication No. 2003/0188035 A1; and

(8) U.S. Patent Publication No. 2003/0188229 A1.

(e) Set forth below is a detailed discussion of references which points out with particularity how the claimed subject matter is distinguishable over the references.

A. Claimed Embodiments of the Present Invention

The claimed embodiments relate to a storage system, more particularly to a distributed storage system configured to perform a remote copy function. Primary storage subsystems synchronously send write data to an intermediate storage system. Upon receiving the write data, the intermediate storage system generates control data, which contains information to identifying the write order. A portion of the control data, e.g., a sequence

number, is generated and attached to the control data of the write data. The intermediate storage system sends the write data along with its control data to the secondary storage systems asynchronously. The secondary storage systems store the write data to the secondary volumes based on the control data. The write order is maintained using the sequence number generated by the intermediate subsystem.

Independent claim 1 recites a remote copy system, comprising first and second primary storage subsystems, the first primary storage subsystem including a first primary volume, the second primary storage subsystem including a second primary volume, the first and second primary volumes storing a plurality of write data in a given order. An intermediate storage subsystem is configured to receive the write data from the first and second primary storage subsystems. The intermediate storage subsystem includes a write-order-information provider that is configured to generate write-order information for the write data received from the first and second primary storage subsystems, the write order information being associated with the write data received from the first and second primary storage subsystems, the write order information reflecting the given order of storage of the write data in the first and second primary storage subsystems. First and second secondary storage subsystems are configured to receive the write data from the intermediate storage subsystem, the first secondary storage subsystem including a first secondary volume that is configured to mirror the first primary volume, the second secondary storage subsystem including a second secondary volume that is configured to mirror the second primary volume, wherein the write data are stored in the first and second secondary storage subsystems according to the write order information associated with the write data.

Independent claim 10 recites an intermediate storage subsystem provided in a remote copy system and coupled to a plurality of primary storage subsystems and a plurality of secondary subsystems. The intermediate storage subsystem comprises a first storage area configured to receive write data from at least one primary subsystem, the write data being received synchronously from the at least one primary subsystem; and a write-order-information provider configured to generate write order information for the write data received from the at least one primary subsystem, the write order information being associated with the write data. The write order information is used to store the write data in

at least one of the secondary subsystems, so that the at least one secondary subsystem mirrors the at least one primary subsystem.

Independent claim 19 recites a method for operating a remote copy system. The method comprises receiving first write data from a first primary storage subsystem at an intermediate storage subsystem, the first write data being sent by the first primary subsystem synchronously; associating first write order information to the first write data; receiving second write data from a second primary storage subsystem at the intermediate subsystem, the second write data being sent by the second primary subsystem synchronously; associating second write order information to the second write data; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem. The first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information.

Independent claim 22 recites a computer readable medium comprising a computer program for operating a remote copy system. The computer program comprises code receiving first write data from a first primary volume of a first primary storage subsystem at an intermediate storage subsystem, the first write data being sent synchronously by the first primary subsystem; code for associating first write order information to the first write data; code for receiving second write data from a second primary volume of a second primary storage subsystem at the intermediate subsystem, the second write data being sent synchronously by the second primary subsystem; code for associating second write order information to the second write data; code for transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem, the first secondary subsystem including a first secondary volume; and code for transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, the second secondary subsystem including a second secondary volume. The first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, so that the first and second secondary volumes mirror the first and second primary volumes.

Independent claim 23 recites an intermediate storage subsystem provided in a distributed remote copy system. The intermediate storage subsystem comprises means for receiving write data from first and second primary volumes of first and second primary subsystems, the first primary volume being defined in the first primary subsystem, the second primary volume being defined in the second primary subsystem, the write data being received synchronously from the primary subsystems; and means for generating write order information for the write data received from the primary subsystems, the write order information being associated with the write data, the write order information providing information as to a write order of the write data. The write order information is used to store the write data in the first and second secondary volumes of first and second secondary subsystems, the first secondary volume being defined in the first secondary subsystem, the second secondary volume being defined in the second secondary subsystem. The first and second secondary volumes mirror the first and second primary volumes.

One of the benefits that may be derived is that it provides a data storage system or remote copy system that provides the benefits of the synchronous and asynchronous modes, i.e., enables the primary and secondary systems to be placed far apart while guaranteeing no data loss.

B. Discussion of the References

1. U.S. Patent No. 5,446,871

The patent to Shomler et al., US 5446871, discloses a method and system for asynchronous remote data duplexing at a distant location from copies based at a primary site storage subsystem 5, which copying is non-disruptive to executing applications, and further in which any data loss occasioned by losses in flight or updates never received at the time of any interruption between the primary 1 and remote sites 9 are accounted for at the remote site. The method assigns a token and unique sequence number responsive to each write operation at the primary site, and sends the tokens+numbers and data updates to the remote site. The method relies upon the sequence number to establish a sequency and define gaps therein to ascertain missing updates. More specifically, the serializer portion of DSM assigns a write sequence token to every write operation and puts these tokens into messages for VTAM 11 to send to a receiving system (VTAM 15 and Data Mover 17) at the secondary

location (remote site 9). Also, the data mover portion of DSM 13 obtains changed data-records written to the primary DASD-and forms them and their tokens into messages for VTAM 11 to send to the secondary site 9. See column 6, lines 29-37.

Although Shomler et al. discloses the use of a sequence number to establish sequency, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Shomler et al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

2. U.S. Patent No. 5,504,861

The patent to Crockett et al., US 5504861, discloses a remote data shadowing system that provides storage based, real time disaster recovery capability. Record updates at a primary site 14 cause write I/O operations in a storage subsystem therein. The self describing record sets are transmitted to a remote secondary site 15 wherein consistency groups are formed such that the record updates are ordered so that the record updates can be shadowed in an order consistent with the order the record updates cause write I/O operations at the primary site. See column 12, lines 47-50. The record updates are written according to full consistency group recovery rules such that should the primary site be unavailable, the secondary site can recover a consistency group. See column 17, lines 31-47. The asynchronous data shadowing system 400 encompasses collecting control data from the

primary storage controllers 405 so that an order of all data writes to the primary DASDs 406 is preserved and applied to the secondary DASDs (preserving the data write order across all primary storage subsystems). The data and control information transmitted to the secondary site 431, must be sufficient such that the presence of the primary site 421 is no longer required to preserve data integrity. See column 10, lines 45-53.

Crockett et al. is directed to remote data duplexing. Although it discloses record updates and consistency in recovery to preserve an order of data writes, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Crockett et al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

3. U.S. Patent No. 5,734,818

The patent to Kern et al., US 5734818, discloses forming consistency groups using self-describing record sets for remote data duplexing. A remote data shadowing system provides storage based, real time disaster recovery capability. Record updates at a primary site cause write I/O operations in a storage subsystem therein. The write I/O operations are time stamped and the time, sequence, and physical locations of the record updates are collected in a primary data mover (PDM). The primary data mover groups sets of the record updates and associated control information based upon a predetermined time interval, the

primary data mover appending a prefix header to the record (updates thereby forming self describing record sets. The self describing record sets are transmitted to a remote secondary site wherein consistency groups are formed such that the record updates are ordered so that the record updates can be shadowed in an order consistent with the order the record updates cause write I/O operations at the primary site. FIGS. 5 and 6 show a journal record format created by the PDM 404 for each self describing record. A time interval group number 503 is supplied by the PDM 404 to identify a time interval for which the current record sets belong. The operational time stamp 502 and the records read time 507 are used by the PDM 404 to group sets of read record sets from each of the primary storage controllers 405. The record set information 600 is generated by the primary storage controllers 405 and collected by the PDM 404. The update records are handled in software groups called consistency groups so that the secondary data mover (SDM) 414 can copy the record updates in the same order they were written at the primary DASDs 406. After all read record sets across all primary storage controllers 405 for a predetermined time interval are received at the secondary site 431, the SDM 414 interprets the received control information and applies the received read record sets to the secondary DASDs 416 in groups of record updates such that the record updates are applied in the same sequence that those record updates were originally written on the primary DASDs 406. Thus, all primary application order consistency is maintained at the secondary site 431. See column 11, line 14 to column 12, line 67.

Although Kern et al. discloses the use of consistency groups to preserve the sequence of record updates, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Kern et al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the

first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

4. U.S. Patent No. 6,658,540 B1

The patent to Sicola et al., US 6658540, discloses a method for transaction command ordering in a remote data replication system. A disaster-tolerant data backup and remote copy system 100 which is implemented as a controller-based replication of one or more LUNs (logical units) between two remotely separated pairs of array controllers 201/202 and 211/212 connected by redundant links. The system provides a method for allowing a large number of commands to be outstanding in transit between local and remote sites while ensuring the proper ordering of commands on remote media during asynchronous or synchronous data replication. In addition, the system provides a mechanism for automatic `tuning` of links based on the distance between the array controllers. FIG. 14 is a flowchart illustrating an exemplary method used by the system to ensure proper write ordering on the remote media. At step 1405, host computer 101 issues a write command to initiator controller A1 (see FIG. 3). At step 1410, each write command is assigned a sequence number when it is received by the controller; i.e., the Command Identifier for the present command is set to the current Sequence Number (SQN). The initial sequence number has a value of 1, and subsequent sequence numbers are generated by incrementing a 16 bit sequence number counter. Accordingly, the current SQN is incremented at this point. At step 1412, controller A1 receives write data from host computer 101. Next, at step 1415, when a command (i.e., the write to the local storage array) completes, the current value of the sequence number SQN is also stored in the command's control block. The system 100 merges the write commands in the proper order. See column 17, line 38-56; column 18, lines 35-51.

Although Sicola et al. discloses the use of a sequence number to ensure the write commands are merged in the proper order, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Sicola et

al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

5. U.S. Patent Publication No. 2003/0014432 A1

The published patent application of Teloh et al., US 20030014432, discloses a method and apparatus for performing remote data replication. The data replication facility is able to group together structures (such as a write ahead log and a corresponding table entry) into a single data set while preserving the write ordering for each asynchronous writer. In this manner, two separate processes or threads can run asynchronous to each other and can copy or mirror their respective volumes to remote storage devices while preserving their respective write order. Moreover, an application utilizing multiple volumes for write order sensitive data can be replicated as a group or single entity while preserving the write ordering of the data. FIGS. 6 and 7 illustrate that the illustrative data replication facility is able to replicate a primary volume 100 from the local storage device 24 to multiple mirrored volumes 102 and 104 on one or more remote storage devices. FIG. 6 illustrates the situation in which the multiple mirrored volumes 102 and 104 are located on the same remote storage device 26 while FIG. 7 illustrates the situation in which the multiple mirrored volumes 102 and 104 are located on multiple remote storage devices 26 and 26' respectively. The local data replication facility 20 preserves the write ordering of all volumes copied during the copy operation. See paragraphs [0053] and [0056]-[0058].

Although Teloh et al. discloses a technique to preserve the write ordering of volumes copied, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Teloh et al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

6. U.S. Patent Publication No. 2003/0074378 A1

The published patent application of Midgley et al., US 20030074378, discloses systems and methods for backing up data files. The invention provides systems and methods for continuous back up of data stored on a computer network. To this end the systems of the invention include a synchronization process that replicates selected source data files data stored on the network and to create a corresponding set of replicated data files, called the target data files, that are stored on a back up server. A dynamic replication process includes a plurality of agents, each of which monitors a portion of the source data files to detect and capture, at the byte-level, changes to the source data files. The dynamic replication process may include a write order controller that is responsive to the time stamp signal for controlling the order in which recorded changes are written to the target data file. The method controls the order in which changes are written to the target data files, thereby ensuring that in case of an interruption in service, the target data file will have been amended to correspond to an actual version of the source data file. As shown in FIG. 6, the agent

process 30 can detect that a journal file contains information and can transfer the journal file to the backup server 12. The backup system may process the journal file to identify the time stamp information and to ensure that changes and modifications made to a target data file occur in the write order sequence of the corresponding source data file. See paragraphs [0063]-[0065].

Although Midgley et al. discloses the use of a journal file with time stamp information that can be used to preserve the write order sequence, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Midgley et al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

7. U.S. Patent Publication No. 2003/0188035 A1

The published patent application of Lubbers et al., US 20030188035, discloses a system for communicating between two devices in a network in which a semi-persistent tunnel is established between the two devices in advance of data communication. Data replication management (DRM) groups 505 comprise a set of related virtual disks or LUNs that belong to copy sets all of which have the same source and destination. DRM groups 505 are used for maintaining crash consistency and preserving write ordering. See paragraph [0060]. An important feature of the data transfer protocol is that it enables a group to

maintain write ordering among the members. To ensure write order preservation, a group sequence number (GSN) is associated with each write operation. Each write operation in a stream of operations has a unique GSN value that identifies the order in which that write operation was completely received in relation to all other write operations for that group. The set of GSNs for a stream of operations forms an unbroken, continuous sequence of values with no gaps. The GSN is stored atomically with the write operation so that a write operation will not be valid until a GSN is associated with the operation. See paragraph [0080].

Although Lubbers et al. discloses the use of a group sequence number (GSN) to preserve write ordering, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Lubbers et al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

8. U.S. Patent Publication No. 2003/0188229 A1

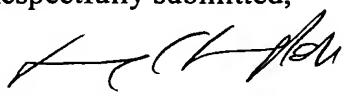
The published patent application of Lubbers et al., US 20030188229, discloses a system and method for managing data logging memory in a storage area network. A data storage system adapted to maintain redundant data storage sets at a destination location(s) is disclosed. Data replication management (DRM) groups 505 comprise a set of related virtual disks or LUNs that belong to copy sets all of which have the same source and destination. DRM groups 505 are used for maintaining crash consistency and preserving write ordering.

See paragraph [0059]. A group maintains write ordering among the members for asynchronous operation and logging/merging. During the time delay of copying in an asynchronous operation, the various replicas are inexact. When asynchronous operation is allowed, it is important that all replicas eventually implement the modification. Since multiple modification operations may be pending but uncommitted against a particular replica, it is necessary that the original order in which the modifications were presented be preserved when the pending modifications are applied to each replica. To ensure write order preservation, a log is maintained for each group 705 that records the history of write commands and data from a host. An ordering algorithm uses a group sequence number (GSN) and the remote groups 705 ensure that the data is written in order sequence. See paragraphs [0060]-[0061].

Although Lubbers et al. discloses the use of a log and a group sequence number to preserve write ordering, it does not disclose an intermediate storage subsystem that generates write-order information for the write data received from a primary storage subsystem and sends it to a secondary storage subsystem. More specifically, Lubbers et al. fails to teach an intermediate storage subsystem including a write-order-information provider or means for generating write-order information for the write data received from one or more primary storage subsystems, the write order information being associated with the write data, wherein the data are stored in one or more secondary storage subsystems according to the write order information associated with the write data, as recited in independent claims 1, 10, and 23; or associating first write order information to the first write data from a first primary storage subsystem; associating second write order information to the second write data from a second primary storage subsystem; transmitting asynchronously the first write data and the first write order information to a first secondary storage subsystem; and transmitting asynchronously the second write data and the second write order information to a second secondary storage subsystem, wherein the first and second write data are stored in the first and second secondary subsystems, respectively, according to the first and second write order information, as recited in independent claims 19 and 22.

(f) In view of this petition, the Examiner is respectfully requested to issue a first Office Action at an early date.

Respectfully submitted,



Chun-Pok Leung
Reg. No. 41,405

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 650-326-2400
Fax: 415-576-0300
Attachments
RL:rl
60601036 v1